Unsupervised parsing of gaze data with a beta-process vector auto-regressive hidden Markov model

Joseph W. Houpt¹ · Mary E. Frame² · Leslie M. Blaha³

Published online: 26 October 2017
© Psychonomic Society, Inc. 2017

Abstract The first stage of analyzing eye-tracking data is commonly to code the data into sequences of fixations and saccades. This process is usually automated using simple, predetermined rules for classifying ranges of the time series into events, such as “if the dispersion of gaze samples is lower than a particular threshold, then code as a fixation; otherwise code as a saccade.” More recent approaches incorporate additional eye-movement categories in automated parsing algorithms by using time-varying, data-driven thresholds. We describe an alternative approach using the beta-process vector auto-regressive hidden Markov model (BP-AR-HMM). The BP-AR-HMM offers two main advantages over existing frameworks. First, it provides a statistical model for eye-movement classification rather than a single estimate. Second, the BP-AR-HMM uses a latent

process to model the number and nature of the types of eye movements and hence is not constrained to predetermined categories. We applied the BP-AR-HMM both to high-sampling rate gaze data from Andersson et al. (*Behavior Research Methods* 49(2), 1–22 2016) and to low-sampling rate data from the DIEM project (Mital et al., *Cognitive Computation* 3(1), 5–24 2011). Driven by the data properties, the BP-AR-HMM identified over five categories of movements, some which clearly mapped on to fixations and saccades, and others potentially captured post-saccadic oscillations, smooth pursuit, and various recording errors. The BP-AR-HMM serves as an effective algorithm for data-driven event parsing alone or as an initial step in exploring the characteristics of gaze data sets.

Keywords Eye tracking · Fixation · Saccade · Smooth pursuit · Indian buffet · Vector auto-regressive · Hidden Markov model

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-017-0974-7>) contains supplementary material, which is available to authorized users.

✉ Joseph W. Houpt
joseph.houpt@wright.edu

Mary E. Frame
mary.frame@wright.edu

Leslie M. Blaha
leslie.blaha@pnnl.gov

¹ Department of Psychology, Wright State University, Dayton, OH 45435, USA

² Department of Psychology, Miami University, Oxford, OH 45056, USA

³ Visual Analytics, Pacific Northwest National Laboratory, Richland, WA 99354, USA

Unsupervised parsing of gaze data with a beta-process vector auto-regressive hidden Markov model

The first step in nearly every application of eye tracking is segmenting the gaze positions into sequences of movement types, such as fixations and saccades. These segments are then available for interpretation in terms of cognitive task manipulations, natural behavior strategies, or clinically diagnostic behaviors, depending on the research motivating the data collection. The depth and quality of inferences drawn from these segments are clearly affected by the quality of the gaze segmentation process and the assumptions underlying that process. We propose that initial eye-tracking data parsing can be augmented using an unsupervised

learning approach in which the ocular dynamics represented in the data drive the event identification process. Unsupervised parsing offers an alternative, exploratory data analysis step that can occur in parallel with or as an alternative to a more traditional initial gaze segmentation process. The latter might be thought of as parsing driven by the assumptions of the investigators about movement characteristics, where the unsupervised parsing is driven by the data characteristics. In this paper, we demonstrate the use of a state-of-the-art machine learning algorithm, based on the combination of a beta process, a vector auto-regressive process, and a hidden Markov model (henceforth abbreviated as BP-AR-HMM) developed by Fox and colleagues (Fox, 2009; Fox, Sudderth, Jordan, & Willsky, 2011; Fox, Hughes, Sudderth, & Jordan, 2014).

The initial focus of gaze segmentation algorithms was accurately defining fixations. This was because eye tracking was first established as a tool for assessing attention in experimental paradigms with static images or in studies of reading. An early and influential approach proposed by Mason (1976) is the delta method, which uses a series of logic gates for parsing fixations from the time series of raw (x, y) screen/image positions. In this approach, a fixation is defined by positional dispersion, i.e., the amount of position displacement between subsequent samples. Sequential movements contained within an experimenter-determined dispersion threshold are labeled a fixation event. In many applications, fixation identification is sufficient, as fixations are overt markers of a viewer's locus of attention.

However, as eye tracking increases in popularity and diversity of application, including more dynamic and naturalistic stimuli, non-fixation events have become more important in analysis (e.g., Duchowski, 2002; Engbert & Kliegl, 2003; Hayhoe & Ballard, 2005; Rayner, 1998). Eye-tracking algorithms have consequently evolved into more multifaceted classifications of eye movements, adding identification of saccades and other behaviors to the identification of fixations.

Several types of eye movements are of potential interest in modern analysis of eye-tracking data. Based on a combination of the musculature of the eye, neural control mechanisms, and the viewing conditions, six primary types of eye movements have been identified: fixations, saccades, smooth pursuit, optokinetic reflex, vestibulo-ocular reflex, and vergence (e.g., Karpov & Komogortsev, 2011; Land & Tatler, 2009). In some studies, further segmentation of the behaviors, particularly within-fixation movements, has been of interest. For example, researchers have examined component movements of fixations such as microsaccades, tremor or nystagmus, and drift (e.g., Engbert & Kliegl, 2003; Otero-Millan, Castro, Macknik, & Martinez-Conde, 2014). Particularly in dynamic viewing conditions, saccadic activity may raise questions about several (at least seven) related

or component movements, including pre- and post-saccadic oscillations (dynamic overshoots) and glissadic movements (Larsson, Nyström & Stridh, 2010; Nyström & Holmqvist, 2013; see also Bahill & Troost, 1979, for the interrelationships of different saccadic movement types). Blinks and recording errors must also be accounted for, particularly in cases when blinks create artifacts that an automated parser might interpret as saccades.

Although classification algorithms have attempted to provide accurate parsing of eye-movement data into saccades, glissades, fixations, and smooth pursuit, most of these algorithms have relied on predetermined but inconsistent thresholds or definitions for sample classification (Nyström & Holmqvist, 2010; Kapoula, Robinson, & Hain, 1986). Rather than relying on these inconsistent threshold criteria, we propose the use of a data-driven approach to gaze parsing, which allows us to query the data for the different types of states to which gaze samples should be assigned. It also uses the movement dynamics in the recording to define the possible state classifications. By using an unsupervised learning technique, we are able to find criteria for different types of eye movements from data, rather than imposing strict thresholds for categorization. While this approach may not offer a complete off-the-shelf solution to gaze parsing, the data-driven nature means that it offers more flexibility and a reduction of researcher degrees of freedom (Gelman & Loken, 2014; Simmons, Nelson, & Simonsohn, 2011) than hand-picked thresholds.

A potential drawback of an unsupervised technique is that the resulting classifications may not agree with established categories of eye movements. We argue that, as long as the approach can discriminate the most well-agreed-upon categories, fixations and saccades, additional categories identified by the algorithm are indicative of important trends in the data. Another potential pitfall is that the algorithm could segment the data into too many categories. While there is some divergence in the literature with respect the granularity of categories (for example some approaches parse micro-saccades from fixations, Otero-Millan, Troncoso, Macknik, Serrano-Pedraza, & Martinez-Conde, 2008), there is certainly an upper limit on the number of categories that are useful to consider.

Another issue that arises in eye-movement parsing is the extent to which category properties may vary across tasks and individuals. For example, an appropriate upper limit on eye-movement speed in a reading task may lead to problematic segmentation when used with a dynamic scene viewing task. One of the most powerful features of the BP-AR-HMM is that it both allows for states corresponding to particular behaviors to be shared across time series and it allows for unique states to be used. Hence, the model allows for a “fixation” to have the same properties between two participants but does not require it.

In the remainder of the introduction, we give an overview of existing techniques used to parse gaze sequences. This review is not meant to be comprehensive, but rather to highlight the main algorithmic and parametric structures that are normally used. Salvucci and Goldberg (2000) defined a basic taxonomy for parsing algorithms based on the measures they emphasized: velocity-based, dispersion-based, and area-based algorithms. In the present work, we will focus on the first two categories, as our work emphasizes the use of eye-movement dynamics in a manner agnostic to the task or stimulus characteristics. Next, we give a more detailed overview of the BP-AR-HMM approach. Then we will demonstrate the BR-AR-HMM on an available data set previously used to compare other parsing algorithms, as well as on a second open source data set viewing various dynamic stimuli.

Trends in eye-movement parsing algorithms

Static threshold approaches

The majority of eye-parsing algorithms derive from the delta method by Mason (1976), using pre-defined thresholds for event categorization. The delta method evolved into the I-DT or identification by dispersion threshold approach, attributed to Widdel (1984). In I-DT, a maximum dispersion threshold is set such that fixations are sequences of points with dispersion less than the threshold. A second fixed minimum duration threshold is used such that fixations must first stay within the maximum spatial area but for a minimum length of time. I-DT effectively parses fixations, but must be paired with additional saccade detecting algorithms to go beyond fixation identification.

One early but still popular approach, the heuristic filtering approach, adds velocity estimates to the basic delta approach to parse fixations and saccades (Stampe, 1993). This method underlies the default algorithm in some commercial systems, including the EyeLink eye tracker (SR Research, 2007). Stampe (1993) attempted to encode logical rules used by human researchers to encode data: that fixations are periods of little or no motion, and saccades are periods of rapid motion or jumps in position. Filters are used in this algorithm to detect non-monotonicities, so noise is smoothed out of the data. Very short saccades (which might now be classified as micro-saccades) are classified as noise within a fixation.

Because the EyeLink algorithm is popular in basic applications of eye-tracking technology, we next review the algorithm in detail. We note that EyeLink was used to parse the low-sampling rate data that we use in our second BP-AR-HMM application. The EyeLink classifications become a ground truth against which we can understand the new BP-AR-HMM's performance. The EyeLink algorithm incorporates three static threshold algorithms to define saccades:

a velocity-based threshold, an acceleration-based threshold, and a dispersion threshold for saccades (SR Research, 2007). To be classified as a saccade using the velocity threshold, there must be a velocity of 22 °/s or greater, with a higher velocity threshold lengthening the duration threshold for fixations. Using this criterion, this detects saccades as small as 0.3°, which is ideal for smooth pursuit. The EyeLink default acceleration threshold defines saccades as 4,000 °/s² for small saccade detection and 8,000 °/s² for cognitive research. Finally, using a dispersion-based method of classification, appropriate dispersion thresholds are .1–.2° for short saccades and .4–.5° for long saccades.

Although static threshold algorithms are an effective first pass at data classification, there are notable limitations to using static thresholds. Static thresholds are extremely sensitive to user-set parameters. Shic, Scassellati, and Chawarska (2008) found that changing the parameter for fixation dispersion from 1.5° to 3° of visual angle caused the mean fixation duration for one type of stimulus class (either full color photographs or color block arrays) to be greater or lower than the mean fixation duration of the alternative stimulus class. Although adjustment of static thresholds may not always lead to something as dramatic as completely reversing the conclusions of an experimental analysis, static thresholds still might not be optimal for classifying eye movements in all experiments. This may be especially problematic for classifying eye movements of viewers who show more data variability, such as infants or clinical populations.

Iterative static threshold approaches

With many of the static threshold approaches, certain eye-movement events are essentially determined through a process of elimination by parsing data one category at a time. The types of eye-movement events are treated as mutually exclusive categories. After a sample is assigned to an event type, the labeling is not necessarily revisited. Consequently, events may be assigned to one category early that might better fit a subsequent event type (see also Andersson, Larsson, Holmqvist, Stridh & Nyström, 2016). Some algorithms attempt to address this with iterative approaches that make multiple passes through the data to re-check event class assignments, with a goal of a more parsimonious event classification result.

MarkEye is an example of an iterative threshold algorithm that tags fixation, saccade, and smooth pursuit events, but does not separately classify glissades or micro-saccadic eye movements (Berg, Boehnke, Marino, Munoz, & Itti, 2009; Smith & Mital, 2013). This algorithm tags candidate saccade time-points using a two-step threshold-based process. First, it creates a smooth estimate of gaze position. The smoothing is done with a low-pass 63 Hz Butterworth filter

over the raw (x, y) gaze position data. Those positions with velocity greater than 30 %/s are labeled as candidate saccades. The directionality of movement is then computed with a principal component analysis (PCA). A sliding window PCA is computed from the smoothed gaze positions, and the ratio of the second principal component (PC) to the first PC is extracted to indicate relative movement in two directions. The algorithm computes PCAs from several different window sizes, and the resulting ratios are linearly combined for a smooth, robust estimate. The PC ratios of the candidate saccades are checked for constant directionality. If any gaze points are shown to have PC ratios ≈ 0 but insufficient velocity to be a saccade, they are deemed smooth pursuit or other (e.g., blink) events. All remaining data points are labeled fixations. At this point, the algorithm iterates back over the saccade events to collapse nearby movements together in two ways. If two saccades are separated by either a fixation event of short duration (less than 60–80 ms) or a smooth pursuit event with only minor dispersion in the direction of movement (less than 45°), then those saccades are collapsed into a single saccade. If a movement event has small amplitude of less than $1\text{--}2^\circ$ or short duration less than 15–20 ms, it is collapsed into adjacent movements. This final combination works to reduce the rate of false positives among saccade events.

An advantage of MarkEye is that it combines position and velocity information and leverages statistical analysis at different time window lengths to produce a more robust estimation of the classification. However, it still relies on experimenter-determined thresholds for the saccadic activity. This can result in experimenters using different threshold values, as is the case looking between Berg, Boehnke, Marino, Munoz, and Itti (2009) and Smith and Mital (2013) who report using different dispersion and duration criteria. This suggests either different criteria between laboratories, or the measured characteristics of eye movements depend on stimulus, experiment, or task conditions. And although not discussed in the earlier papers, the results of this algorithm will also be sensitive to the filter properties used in smoothing the data. However, MarkEye has seen some strong successes parsing more naturalistic stimulus viewing in both human and primate observers (Berg et al., 2009). This suggests that eye movements elicited by combinations of static and dynamic stimuli will benefit from iterative processing and refining of the event class assignments based on the decomposition of movement dynamics.

In a different approach to iteratively parsing eye movements, Salvucci and Anderson (1998, see also Komogortsev, Gobert, Jayarathna, Koh, & Gowda, 2010) introduced a method leveraging cognitive process models, namely ACT-R, combined with hidden Markov models, to trace the saccade and fixation events in a task-related gaze recording. Cognitive process models simulate task performance,

including related eye movements. They offer a way to decompose a task into the supporting cognitive states. Under the reasonable assumption that different cognitive states relate to eye movements in meaningful ways, then modeling the cognitive states can aid in the interpretation of related gaze events. Salvucci and Anderson's hidden Markov model identification (I-HMM) uses a two-state HMM to provide a stochastic representation of the gaze events. The HMM states represent the velocity distributions for fixations and saccades. The transition probabilities in HMM represent the probabilities of staying in either the fixation or saccade state and the probabilities of transitioning between them. In Salvucci and Anderson (1998), the HMM is then combined with a cognitive process model to optimize transition probabilities and state assignments using both cognitive task demands and the eye-tracking dynamics. In application, I-HMM starts by estimating the velocity of the eye movements. The second step uses a Viterbi sampler (Forney, 1973) to classify each point as a fixation or saccade based on the probabilistic parameters of the initial HMM state, the transition probabilities, and the observed probability distributions derived from the process model. A Baum-Welch re-estimation algorithm (Baum, Petrie, Soules, & Weiss, 1970) is then iteratively applied to optimize the state assignments. The I-HMM approach is more robust than static threshold approaches to changes in dynamic characteristics, but it is limited in its present form to parsing only fixations and saccades. However, because it leverages hidden states and cognitive models, I-HMM has some advantages that we will return to in the discussion.

Adaptive threshold approaches

Adaptive algorithms can adjust threshold parameterizations to better capture variability in gaze data due to individual participant differences or to eye-tracker system/recording settings. Thresholds can be allowed to adapt between individuals or even within a single time series to changes in the local properties of the gaze dispersion.

Goldberg and Schryver (1995) offer an early example of an adaptive gaze parsing technique focused on gaze location. In their analyses, gaze locations are grouped based on proximity using Prim's algorithm for identifying the minimum spanning tree. In particular, a gaze location is selected as the initial point. Then the additional gaze locations are added to the closest node in the tree until all samples are included in the tree. Once the tree is formed, clusters are determined based on the relative distance between gaze location among neighbors while accounting for local vertex density. By using a criterion of relative distance rather than a fixed distance, fixation-like clusters can be formed with different effective dispersion criteria. One major disadvantage of this approach for our purposes is that it ignores

the temporal structure. Hence, the algorithm will group two fixations together if they occur in the same region, even if they are separated temporally by gaze movement to another location.

Nyström and Holmqvist's (2010) method for characterizing eye-tracking data utilizes an adaptive velocity-based algorithm to segment saccades, fixations, and glissades. In their analysis, thresholds for saccade detection are determined through an iterative procedure. The algorithm is initialized with a velocity between 100–300 °/s. After identifying time-segments above the threshold, the window associated with each saccade is expanded up to the first local minimum velocity prior to the saccadic peak (saccade onset threshold) and the first local minimum following the peak (saccade offset threshold). Once saccades are identified in this manner, the mean and standard deviation of the remaining velocity samples are used to calculate a new threshold (specifically 6 standard deviations above the mean of the non-saccade samples). The procedure is then repeated with the new saccade threshold.

Nyström and Holmqvist's algorithm identifies glissades as time-windows during which velocity is greater than the saccadic onset threshold following a fixed delay from the last temporal boundary of a saccade. For what they termed "high-velocity" glissades, the criteria for classification are that the event's velocity curve occurs within a 40 ms window following a classified saccade and that the curve rises above the peak saccade threshold and down below it at least once. These criteria make such events otherwise classifiable as saccades but shorter in duration. In contrast, the criterion for low-velocity glissades requires only that the velocity curve rises above the saccade offset threshold. For both types of glissades, the onset is defined as the offset of the previous saccade.

Fixations are gaze activities classified as neither saccades, glissades, or noise that lasts longer than a set fixation duration threshold of 40 ms. In this algorithm, noise is defined either as unclassified samples that are under the minimum fixation duration threshold or as blinks.

A benefit of the Nyström and Holmqvist type of model is that it does not rely on the pre-selected parameterizations of eye-tracking software for saccadic thresholds and is, instead, more data-driven. It is also less sensitive to noise and drifts in data compared to dispersion threshold-based algorithms (e.g., I-DT). Using velocity rather than position to characterize saccades also provides more accurate temporal information about saccade onsets and offsets.

Along a similar vein, other recent algorithms have proposed to leverage statistics about local properties or sliding windows around samples to allow algorithm parameters to adapt to variability in the data. Larsson, Nyström, and Stridh (2013) uses adaptive local statistics about movement accel-

erations and a velocity-based filter to tease apart saccades and post-saccadic oscillations in the presence of smooth pursuit. Because it emphasized the saccade onset and offset detection, it does not address fixations or the identification of smooth pursuit itself. van der Lans, Wedel, and Pieters (2011) introduced an adaptive algorithm, called the Binocular-Individual Threshold (BIT) algorithm, for fixation identification that uses the covariance in dispersion of binocular eye-tracking recordings. Their approach also uses a sliding window of velocity estimates, which adapts to the data variability within the window. BIT is unique because of the emphasis on binocular data, rather than the monocular estimates used in most algorithms. In our application of the BP-AR-HMM, we use monocular gaze sequences, but it is straightforward to adapt the approach for binocular data.

Functional approaches

A small set of available gaze parsing algorithms rely on functional data estimates, wherein some estimate is made of either the whole time series or some time window within the time series. This is in contrast to many of the aforementioned algorithms that rely on a single statistic or point estimate in the evaluation process.

Sauter, Martin, Di Renzo, and Vomscheid (1991) used a Kalman filter approach to quantify the probability that a segment of eye-tracking data is a saccade movement. Probability was estimated under the null hypothesis that a segment is smooth pursuit, which is characterized as a white noise process simulated with the Kalman filter. Saccades are defined as "jumps" in the mean of the differentiated signal, which is an estimate of the velocity derived from the raw (x , y) position time series. Statistically, they represent this differentiated signal with an auto-regressive moving average (ARMA) process, or an AR process, assuming the velocity signal is corrupted by white noise. Then the difference between the baseline Kalman filter model and the ARMA (or AR) models is estimated, such that this estimator is maximal under the null hypothesis and minimal otherwise. Finding a minimum indicates statistical evidence for a saccade. This model does not identify fixations or other movement events beyond estimating the likelihood that an event is a saccade.

This functional approach has similar elements to a threshold model in that eye-movement classifications are made according to some deviation from a null hypothesis estimate. The null hypothesis may be considered similar to a threshold. An important advantage to using functional estimates is that they account for more of the natural, small-scale variability within a time series of gaze data. The parameter estimates are not treated as independent samples, but incorporate correlations and trends over time into

the analysis. Similar to threshold models based on position, velocity, and/or acceleration, functional estimates can also be computed on any of these movement measures.

A commonality between all of the existing algorithms we have reviewed, regardless of their use of static or adaptive thresholds, point or functional estimators, is that they all use an iterative decision tree process for classifying eye-movement events. This is more obvious in earlier papers like Mason (1976, Figures 2 and 3) and Stampe (1993, Figure 2) in which the classification process is illustrated by logical flowcharts. But even the modern adaptive techniques like Nyström and Holmqvist (2010) still make sequential assignments of different movement classes. And most algorithms are still limited to parsing only a small subset of the types of movements that may occur in a single data set. To move beyond hand coding toward algorithm-based parsing, researchers must rely on selecting some combination of algorithms that might work together. The best combinations to use together is not a topic we have found addressed in the literature at all. Rather than rely on the lucky selection of a combination of parsing algorithms, we propose that researchers utilize a data-driven, unsupervised learning approach to parsing eye-movement data, at least as an initial exploratory data analysis step. In this way, the dynamics in the data can inform researchers about the number of potential different classes of movements in a data set. The dynamics can drive the selection of appropriate algorithms for additional parsing and interpretation.

Beta-process vector auto-regressive hidden Markov model

In this section, we give a brief overview of the BP-AR-HMM. For readers interested in more detailed descriptions of the model, including details on model estimation, we suggest (Fox et al., 2011). The BP-AR-HMM model consists of three layers. At the lowest level, the data are modeled as a switching vector AR process, which is similar to Sauter et al. (1991). This structure implies that the position (or speed, or velocity) of the gaze at any given time is a function of previous positions plus some random variation. The data modeled by the AR process could, in general, be multidimensional, such as x and y screen coordinates, or more disparate dimensions such as pupil size and gaze-speed. In our current application of the model, we used the speed and acceleration at each sample of the gaze trajectory as inputs to the model.

Let X_t be a vector representing the gaze at time t and ϵ_t be an independent, identically distributed, zero-mean, normal random variable. In this paper we represent the gaze at each time by the two-dimensional vector of its estimated velocity, V_t , and acceleration, a_t , although the model allows

for other representations, such as the coordinates of the gaze position at that time. Then, we assume for the vector AR model that:

$$X_t = \theta_{0,t} + \theta_{1,t}X_{t-1} + \cdots + \theta_{n,t}X_{t-n} + \epsilon_t.$$

The assumption that it is a *switching* vector AR process means that the θ parameter can change across time. The probabilities for the changes in θ are controlled by the next level in the model.

The middle level of the model assumes a set of discrete states z_k and that the transitions between them follow a Markov process. Because the switching process among parameter values is not directly observed in the gaze sequence, we refer to this level as the hidden Markov model (HMM) layer. Each state is associated with a different mean vector and covariance matrix from which the θ parameters are sampled. Although the θ s are random variables, a state has a fixed collection of θ s across repetitions of that state both within a time series and across time series. Thus, if the process is in state z_k at some time t and again at another time s , in either the same gaze path or another gaze path, then $\theta_{i,t} = \theta_{i,s} = \theta_{i,z_k}$ for $i = 0, 1, \dots, n$. Herein lies the strength of the BP-AR-HMM for gaze sequence parsing: a state associated with a saccade can be modeled with same parameters within and among gaze path sequences processed by the algorithm.

The Markov process assumption means that the state can change from z_k to z_j with some set probability π_{kj} , and that probability does not depend on prior states. Thus, the probability of switching from a “fixation” state to a “saccade” state does not depend on whether the fixation state was preceded by a pursuit state or even how long the process had been in a fixation state. To influence the prior probability of staying in each state longer, i.e., modeling a gaze sequence with behaviors that last longer, the BP-AR-HMM includes a “sticky” parameter which inflates the probability of self-transitions from z_k to z_k for each k .

The highest level of the model is a beta process that determines which states are available for each time series. Some readers may be familiar with this process under the moniker “Indian Buffet Process” following the intuitive description of the process given in Griffiths and Ghahramani (2011). First, a sample from the process can be represented as a matrix with rows indicating time series and columns indicating state. Each cell is either 1, if the state is available for process, or 0 otherwise. The value of each cell is a binomial random variable with the probability parameter specific to the column. Those probability parameters are in turn sampled from a beta process. In principle, this means that a particular state, such as a state associated with smooth pursuit, may be present in one time series (e.g., watching a

video) but not in another time series (e.g., searching a static image). As indicated in the previous paragraph, it is also possible that a state appears across multiple sequences.

Because the beta process does not constrain the possible number of positive probability parameters, there is no limit on the number of states that could be used to describe either an individual time series or a collection of time series. Using this approach, we are not constraining the model to find a fixed number of states, such as saccades, fixations and smooth-pursuit movements, but are instead allowing for any number of states to be identified.

Although a maximum likelihood parameter estimation procedure would be sufficient for some applications, another advantage of the BP-AR-HMM is that the software includes a sophisticated Markov chain Monte Carlo (MCMC) algorithm for Bayesian inference with the model (see the General Application Method section for more on the software). Rather than obtaining a fixed estimate of the number of states, the properties of those states, or which states are shared among any set of time samples, it is possible to estimate distributions over each of those properties.

As mentioned above, another advantage of this approach is that the distinction among movement categories is driven by the data, not by experimenter-defined thresholds that must be determined in advance. In addition to estimating the category distinctions, the model also gives estimates over the number of distinctive categories indicated by the data. This algorithm offers a fair amount of flexibility, including which properties of the data are modeled (e.g., position, velocity, or speed). It also has the ability to identify events across two sequences as the same eye-movement state. Consequently, one person's "fixation" may have the same dynamic properties as another person, or the algorithm can identify the fixation-like behavior as distinct across the sequences. Thereby, the BP-AR-HMM allows for unique gaze dynamics for each person. Rather than assuming all events are the same, as in static threshold algorithms, or events across sequences have distinct thresholds, as in adaptive algorithms, the BP-AR-HMM event definitions are driven by the data. Additionally, BP-AR-HMM harnesses the flexibility of data clustering, like the I-MST algorithm, to determine the appropriate numbers of events together with the properties of those events. Because it estimates all these characteristic simultaneously, we do not introduce any experimenter bias by specifying some decision logic or order in which different event types are identified.

General application methods

To test the BP-AR-HMM on parsing gaze sequences, we applied it to two data sets that differ in their eye-tracker

recording systems and sampling rates. The first data set is a collection from Andersson et al. (2016) of hand-tagged gaze sequences from multiple subjects viewing three types of stimuli: a smooth pursuit task, a static image, and a short video clip. All were sampled at 500 Hz. These data allow us to compare the performance of the BP-AR-HMM to a standard (human expert event coding), rather than simply comparing to another automated parsing algorithm. We refer to this as the Lund dataset.

The second data set is an open-source data set from the DIEM database (Henderson, 2016; Mital et al., 2011). These data include gaze sequences from a large number of participants free-viewing a range of different video clips which are generally longer than the clips included in the first collection. Although the gaze sequences were originally collected at 1000 Hz, the open science data were made available at the frame rate of the stimuli (30 Hz). This data set provides a contrast to the Lund data because the videos are longer and more diverse in the DIEM set, and it allows us to test whether the BP-AR-HMM is robust against a low sampling rate. We review more details about both data sets in the relevant Application sections.

For both applications, the time series input to the BP-AR-HMM were sequences of velocity and acceleration across time. The gaze path of each subject in each condition was input as a separate time series.

Software

Eye-tracking parsing analyses were conducted in MATLAB using the NPBatesHMM Toolbox (Hughes, 2016; Fox et al., 2014; Hughes, Fox, & Sudderth, 2012) and components of the MarkEye algorithm (Berg, Shen, & Itti, 2009). Statistical analyses and visualization of the resultant eye-events data were conducted in R language for statistical computing (R Development Core Team, 2011). To estimate the posterior, 2500 burn-in samples and 2500 additional samples were drawn. Of the 2500 additional samples, every 25th was preserved to reduce auto-correlation leaving 100 samples. Unless otherwise specified, the state used for an eye-tracking sample in the results was the posterior mode state.¹

Preprocessing

Data from both sets were first converted to (x , y) coordinates in degrees of visual angle from their original pixel coordinates. Following the MarkEye algorithm, we

¹ Although 100 posterior samples would be a low number of samples for estimating a full posterior distribution, it was sufficient for estimating the central tendency of the posterior.

estimated the velocity of the eye movements using a combined radial basis function with a discrete derivative taken $x'_t = x_{t+1} - x_{t-1}$, akin to a one-dimensional Sobel filter. These values were then multiplied by the sampling rate so that units were in degrees per second of movement in velocity calculations. To estimate the acceleration, we used a similar combination of a radial basis function and the discrete second derivative, $x''_t = (x_{t+1} - x_t) - (x_t - x_{t-1})$, multiplied by the squared sampling rate.

Data visualization

Eye-movement events of interest are generally characterized well by the degree of dispersion and the maximum velocity, we predominantly visualize groupings of events by their distributions of maximum speed and maximum dispersion. We plot the bivariate dispersion-speed distributions with heat maps, wherein indigo colors are low probability regions and red colors are high probability regions. Univariate histograms are plotted for the maximum dispersion and maximum speed variables alone. These are plotted along the relevant side of the bivariate heat maps, opposite the axes indicating the range of units. Corresponding plots of the maximum speed by acceleration and maximum dispersion by acceleration are available in supplemental materials (https://github.com/jhoupt/BP-AR-HMM_EyeTracking/).

Application 1: BP-AR-HMM applied to high-sampling rate data

Lund data

The first test of the BP-AR-HMM on real-world data was on data from a study by Andersson et al. (2016), as this was collected at a high sampling rate and included hand-coded event classification. This provided us with a robust standard for comparison with our classification. The 34 data files from Andersson et al. (2016) were collected from 17 students at Lund University. Although binocular data were initially collected, only right-eye data was analyzed in Andersson et al. (2016); consequently, we processed only right-eye data with our BP-AR-HMM algorithm. Data were recorded on a Hi-Speed 1250 eye-tracking system from SensoMotoric Instruments GmbH (Teltow, Germany) with the sampling rate at 500 Hz. Participants' heads were stabilized by chin and forehead rests during data collection. Participants were asked to freely view three types of stimuli: 1) static photographs (20 sequences), 2) a linearly moving dot (23 sequences), and 3) real-world videos of moving objects (19 sequences). The videos used in this data set were fairly short clips of relatively simple moving objects, such as a

few seconds of a roller coaster or of cars driving down a road.

Meaningful eye-movement events in six categories were hand-coded by two expert evaluators. For the present work, we used the data encoding by evaluator RA for our ground truth of movement events. We made no modifications to RA's labels. The six event categories coded in the Lund data were: fixation, saccade, post-saccadic oscillation, smooth pursuit, blink, and other. Histograms of maximum dispersion and maximum speed of each movement within each category based on the hand-tagging are shown in Fig. 1. For clarity, we have right-censored the maximum speed and maximum dispersion distributions at the 95th quantile. The Pearson correlations between the right-truncated distributions are reported in the captions of the subfigures. Correlations were tested with a paired sample, two-tailed t-test against the null hypothesis that $r = 0$ and the alternative hypothesis that $r \neq 0$. Results are reported in the subfigure captions in Fig. 1.

The fixation category, Fig. 1a, had relatively symmetrically distributed dispersions across events, mainly between 0.2° and 0.7°, while having highly skewed maximum speeds. Although the majority of fixations had low maximum speeds, a long right tail of the distribution was above 30 °/s. With such a long tail for speed, dispersion and maximum speed were only weakly correlated for fixations. In contrast, both saccades (Fig. 1b) and post-saccadic oscillations (Fig. 1c) had moderate correlations between the maximum speed and maximum dispersion. Saccades were faster and exhibited larger dispersion than other categories, although there was a fair amount of overlap in both speed and dispersion with the other categories. Post-saccadic oscillations have a maximum dispersion distribution similar to fixations but tended to be faster. In contrast, smooth pursuit movements had a similar maximum speed distribution to fixations but tended to have larger dispersions. There was a mild negative correlation between maximum speed and dispersion across smooth pursuit movements.

Among all of the categories hand-tagged in this dataset, there is overlap in the distributions, indicating that a threshold-based algorithm using only speed and dispersion would not sufficiently discriminate among them. Nonetheless, the qualitative trends of each of the categories match established category definitions: fixations are slower and spatially constrained; smooth pursuit is slower but less spatially constrained; saccades are faster and move larger distances; and post-saccadic oscillations are fast, but spatially constrained.

In the next section, we outline the gaze movement states identified by the BP-AR-HMM using similar visualizations and statistics. We compare the data-driven results to hand-tagged event labels.

Fig. 1 Maximum speed-dispersion histograms for each hand-coded category from Andersson et al. (2016). The x -ranges of all histograms are right-censored at the 0.95 quantiles. Sub-captions contain Pearson correlation between truncated max. speed and max. dispersion distributions

Lund data results

Across the 62 sequences of three experimental conditions, the BP-AR-HMM identified nine states in all 100 of the

preserved MCMC samples. States 6 through 9 were almost exclusively identified with errors and undefined samples in the data set, likely either blinks or off-screen samples. We collapse these error states for inclusion in the confusion

matrix, but as the blink and error data do not exhibit non-zero speed and dispersion values, they are left out of the bivariate plots and descriptive statistics. The distributions of events parsed into the different states was approximately 73% in state 1, 15% in state 2, 5% in state 3, 4% in state 4, 2% in state 5, and 1% across states 6–9.

A key difference in results of the BP-AR-HMM from the hand coding is that it parses the time series into a larger number of events in the data. In the original data coding, there was a total of 5,459 hand-classified events. The BP-AR-HMM extracted 22,603 unique events, many of which (2,627) were only a single sample from the eye tracker, lasting 2 ms, which is likely to be too short and small to be a meaningful gaze event. This indicates that further post-processing may be necessary, in which single-sample states are collapsed into neighboring states when appropriate, similar to a stage in the MarkEye algorithm.

In the following analyses, we have censored the data in two ways, to aid in clarifying the mean trends of the dynamics for each state. First, in accordance with the above paragraph, we have left-censored the duration variable of the BP-AR-HMM data such that we are analyzing all events with duration greater than 2 ms. Second, we right-censored the data to remove acceleration outliers within each state. This was done in a three step process: (1) select the data falling into each state; (2) compute the 25th and 75th quantiles for acceleration; and (3) remove any data point with an acceleration outlier defined as greater than 3 times the interquartile range (75th – 25th percentiles), with the exception of state 5 which used 1.5 times the interquartile range. The proportions of data removed as an acceleration outlier was 0.3% in state 1, 0.2% in state 2, 0.4% in state 3, 2.2% in state 4, 1.0% in state 5.

Figure 2 shows sample speed and acceleration traces with color indicating state for each segment. The separate states, particularly the states associated with faster movements and the state associated with small dispersion, seems to match qualitative patterns expected from gaze path data. That is, slow movements indicated in blue are connected by fast movements in shades of yellow and red in Fig. 2.

The bivariate distributions over maximum dispersion and speed for the states extracted by the BP-AR-HMM are illustrated in Fig. 3. The states seem to be described reasonably well by the shapes of the bivariate distribution of maximum speed by maximum dispersion, but we note that the plots of all dispersion, speed, and acceleration combinations are included in the supplemental material. We have organized the plots in Fig. 3 with the states exhibiting the fastest maximum speeds in the top row and states exhibiting the slowest maximum speeds at the bottom. This is reflected by the decreasing ranges of y-axis values. Across

Fig. 2 Velocity by time (top) and acceleration by time (bottom) from the Lund dataset with color indicating BP-AR-HMM state. State 1 is indicated in blue; state 2 is red; state 3 is light red; state 4 is yellow, and state 5 is light yellow

all states, there is some variation in maximum dispersion values, though less variation than in maximum speed. The fastest two states cover a larger range of dispersion values than the slower three states, which all have x -axis values in the range of approximately 0° to 1.2° . Table 1 summarizes the means, standard deviations, and upper and lower 5% quantiles for maximum dispersion, maximum speed, and acceleration values for states 1–5 (using the censored data, as described above).

The slowest state, state 1 in Fig. 3e, has maximum speeds predominantly ranging between approximately $12^\circ/s$ and $30^\circ/s$, with dispersions between 0° and 0.65° . The second slowest state, state 2, displayed in Fig. 3d, ranges between $20^\circ/s$ and $65^\circ/s$ with a range of dispersion similar to the slowest state 1. The next fastest state, state 3 in Fig. 3c, exhibited speeds between approximately $60^\circ/s$ and $190^\circ/s$ while still having a limited dispersion range, below 1.2° . We note that, as shown in Fig. 3, states 1 through 3 all exhibited a moderate positive correlation between maximum dispersion and maximum speed.

The two fastest states, shown in Fig. 3a and b exhibited maximum speeds between $200^\circ/s$ and $950^\circ/s$, and $440^\circ/s$ and $8,500^\circ/s$ respectively, and each had dispersions ranging up to approximately 14° . State 4 exhibited a strong positive correlation between maximum speed and maximum dispersion. Given that saccades speeds over $1,000^\circ/s$ are normally filtered out as biologically implausible, it is likely that the fastest state (state 5), which averaged a maximum speed of nearly $3163^\circ/s$ and a negative correlation with maximum dispersion, represents the transition to or from a blink or some type of recording error. Indeed, only 1.8% of samples

Fig. 3 Maximum speed-dispersion histograms for each BP-AR-HMM non-error state in the Lund data. For clarity in these images only, the depicted dispersion range is right-tail truncated at the 0.95 quantile.

Sub-captions contain the Pearson correlation between max. speed and max. dispersion. All data depicted are censored as described in the text

across all sequences were categorized at this fastest type, and in the confusion matrix, Table 2, the state 5 labels are most often samples originally encoded as undefined/errors.

From the summary statistics in Table 1, there is clear overlap in the distributions of the event characteristics, particularly in their maximum dispersions.

Table 1 Descriptive statistics for the BP-AR-HMM movement states on the Lund data^a

	Mean	Std. Dev.	5% Quantile	95% Quantile
Maximum dispersion (deg.)				
State 1	0.23	0.26	0.03	0.63
State 2	0.24	0.22	0.03	0.59
State 3	0.45	0.35	0.03	1.04
State 4	4.08	3.79	0.30	11.54
State 5	4.71	4.65	0.04	13.67
Maximum speed (deg/sec)				
State 1	21.92	4.02	14.81	27.62
State 2	42.42	11.86	26.32	62.57
State 3	114.87	35.95	66.52	180.55
State 4	486.14	229.06	207.53	928.59
State 5	3,162.90	2,877.92	443.62	8,394.60
Maximum acceleration (deg./sec ²)				
State 1	5,832.17	1,881.48	2,812.44	8,834.21
State 2	7,133.78	2,752.66	2,776.55	11,696.75
State 3	16,476.79	5,674.76	7,016.20	25,138.00
State 4	39,105.87	13,153.40	19,119.90	60,769.50
State 5	335,831.05	320,295.03	67,532.75	917,557.50

^aNote that states 6–9 contain events (blinks and errors) with no meaningful speed or dispersion statistics. These states account for approximately 1% of the total data set

Confusion analysis

Table 2 shows the confusion matrix of event assignments between the hand-coded labels by Andersson et al. (2016) and the results of the BP-AR-HMM process.² State 1 with the slowest movements and lowest dispersion (Fig. 3e) had the most overlap with fixations and smooth pursuit in the hand-tagged data from Andersson et al. (2016). Interestingly, although both smooth pursuit and fixation movements were included, there is no indication of a differentiation between fixation and smooth pursuit (i.e., bimodality), at least with respect to maximum speed and maximum dispersion.

The next slowest state (state 2, Fig. 3d) overlapped mostly with post-saccadic oscillations but also a mix of

Table 2 Confusion matrix of BP-AR-HMM states (columns) against hand-coded events (rows; Andersson et al., 2016)

	State 1	State 2	State 3	State 4	State 5	States 6–9
Fixation	0.862	0.113	0.020	0.003	0.002	0
Saccade	0.005	0.037	0.240	0.641	0.077	0
Post-saccadic oscillation	0.043	0.286	0.528	0.130	0.013	0
Smooth pursuit	0.765	0.195	0.036	0.003	0.001	0
Blink	0	0	0	0	0	1
Undefined	0.297	0.052	0.167	0.253	0.224	0.007

smooth pursuit and fixation events in the hand-tagged data. State 3 also exhibited a tendency to include many post-saccadic oscillation events, as well as a good proportion of saccades. State 4 overlaps mostly with the saccade event state, with a smaller proportion of post-saccadic oscillation events. State 5, which visually appears to be bimodally distributed in Fig. 3a, overlaps predominantly with undefined events in the hand-tagged data.

Based on the relationships between the BP-AR-HMM states and the hand-tagged labels from human experts suggested by Table 2, we explore collapsing the BP-AR-HMM states into fewer grouped states. Because states 2 and 3 similarly reflect the highest overlap with post-saccadic oscillations, we collapsed them into a single state (collapsed grouping 3, Fig. 4c). Note that when we considered collapsing across BP-AR-HMM states, we did not change states 1 and 4. This is because fixations and smooth pursuit were predominantly assigned to state 1, while saccades were generally assigned to state 4. Those states were left as-is, but Fig. 4a refers to state 1 as collapsed grouping 1, and Fig. 4b refers to state 4 as collapsed grouping 2. We plotted the bivariate maximum dispersion-speed distributions of these collapsed groupings in Fig. 4. Given that state 5 had highest association with undefined or error events, we left it out of the collapsing process.

Figure 4 reflects the three dominant trends from the Lund data dynamics in the relationships between maximum speed and maximum dispersion. Like the fixations and smooth pursuit in the original data coding (Fig. 1a and d respectively), ellipsoids of variability show moderate correlation in grouping 1 (mean vector of dynamics $< 0.23^\circ, 22.95^\circ/s, 5,825.47^\circ/s^2 >$; standard deviation vector $< 0.26^\circ, 4.02^\circ/s, 1,874.88^\circ/s^2 >$). However, there is no indication of bimodality in the collapsed grouping 1 (Fig. 4a) that would visually differentiate the two types of movement events. This suggests that dispersion, speed, and acceleration alone are not sufficient for parsing out the smooth pursuit movements using this data-drive approach in *post hoc* analysis.

²Note that we have made no attempt to assign labels to the BP-AR-HMM states. Therefore, the diagonal of Table 2 may not contain the largest probabilities. However, we could re-arrange the table such that the largest proportions of label assignments for each state fall onto the diagonal, as would be seen in a typical identification-confusion matrix. In this case, one might be tempted to provide a typical labeling scheme to these states based on these confusions. To remain data driven and label agnostic, we do not make such assignments. Rather, we encourage the exploration of the dynamics to find dominant patterns, which may or may not correspond to a particular labeling scheme.

Fig. 4 Maximum speed-dispersion histograms for the collapsed groups within the Lund data. The depicted ranges of both dimensions are right-truncated at the 0.95 quantile for clarity in this figure. Sub-captions contain the Pearson correlation between maximum speed and maximum dispersion. All data depicted are censored as described in the text

Additional properties of the behaviors may be necessary to discriminate smooth pursuit and fixations, such as the ratio of PC scores used in MarkEye or any of the other alternatives surveyed in Andersson et al. (2016).

Collapsed groupings 2 and 3 show stronger positive relationships between dimensions. Grouping 2 is similar to the plot of the hand-coded saccade data in Fig. 1b with mean vector (mean vector $< 4.08^\circ$, $486.14.50^\circ/\text{s}$, $39,076.95^\circ/\text{s}^2 >$ and standard deviation vector $< 3.79^\circ$, $228.85^\circ/\text{s}$, $13,142.20^\circ/\text{s}^2 >$. Grouping 3 captures behaviors similar to the hand-coded post-saccadic oscillations, Fig. 1c, though likely also including some saccades and fixations or smooth pursuit events mixed in. The descriptive statistics vectors for collapsed grouping 3 are: mean vector $< 0.38^\circ$, $71.50^\circ/\text{s}$, $11,420.61^\circ/\text{s}^2 >$ and standard deviation vector $< 0.40^\circ$, $48.82^\circ/\text{s}$, $6,889.64^\circ/\text{s}^2 >$. Thus, although we do not want to be too hasty to simply assign labels to the BP-AR-HMM states, we can see similarities to the

expert encodings. More work teasing out the differences in the dynamics that produced the larger number of eye-movement events in the BP-AR-HMM results is needed to really understand the nuances between the parsing results, but that deeper evaluation is beyond the scope of our present application demonstrations.

Application 2: BP-AR-HMM applied to low-sampling rate data

DIEM database

As a follow-up test of the BP-AR-HMM on high-sampling-rate data from Andersson et al. (2016), we ran the algorithm on open-access data files from the DIEM database. These data were down-sampled to a much lower resolution than the Lund data set. The DIEM data also do not include a

hand-tagged parsing of the gaze sequences available for comparison. Instead, the DIEM database includes event tags indicating either saccade, non-saccadic eye movement, or blink, based on EyeLink software's default algorithm (summarized in the Introduction) which was applied to the gaze sequences before downsampling. The goal of applying the BP-AR-HMM to this data set was to determine if the algorithm is robust to lower-resolution data.

Ten videos were selected with 10 subjects per video from the DIEM database (Mital et al., 2011) for the purpose of our analyses. For all of the data in this database, participants were told to simply watch (free view) multiple videos and rate how much they liked each one. Eye-movement recordings were taken on an SR Research EyeLink 2000 desk-mounted eye trackers, initially sampled at 1000 Hz and down-sampled to 30 Hz. The videos selected to be parsed by our algorithm ranged from having a large degree of expected eye motion (i.e., a tennis match) to a low expected range of motion (i.e., a video of interviews with relatively static positions of the interviewees). The specific videos were as follows: 3 tennis matches, 1 amateur ping-pong match, 1 cooking show, 1 movie trailer, 1 video game trailer, 2 interviews, and 1 television show opening credits. The videos varied in length from about 60 to 90 seconds, and the native resolutions were between 860×568 and 1280×720 horizontal by vertical pixels. All videos were viewed on a 21" ViewSonic monitor at a resolution of 1280×960 pixels with a 120 Hz refresh rate. The purpose of selecting videos with variable subject matter was to elicit different types of eye movements from the viewers. The goal of applying BP-AR-HMM to the DIEM data was to determine if the algorithm could correctly parse videos where one would expect primarily saccades and smooth pursuit (tennis match) versus primarily fixations (interview).

The original encoding of the data by Mital, Smith, Hill, and Henderson (2011) used three categories of events:

blinks, saccades, and non-saccadic eye movements. The latter category would have included fixations, smooth pursuit, optokinetic nystagmus, or any other movements that didn't meet the criteria for saccade. Blinks were defined as points where the pupil was missing from the image. Saccades, per the EyeLink algorithm described above, were identified with a 50 °/s velocity threshold and 8,000 °/s² acceleration threshold. All other events were grouped into non-saccadic eye movements.

Descriptive statistics for the two non-blink categories exhibited similar mean tendencies, with non-saccade movements having maximum dispersion-maximum speed mean vector $\langle 0.81^\circ, 119.17^\circ/\text{s} \rangle$ and standard deviation vector $\langle 0.92^\circ, 130.75^\circ/\text{s} \rangle$. The saccade movements have maximum dispersion-maximum speed mean vector $\langle 1.14^\circ, 89.47^\circ/\text{s} \rangle$ and standard deviation vector $\langle 2.46^\circ, 113.13^\circ/\text{s} \rangle$. To plot the data, we note that after down-sampling to 30 Hz (which was done before the data were made open source), single sample events of 33 ms comprised the majority of the saccade data, at 73.1%. Single sample events comprised only 3.0% of the non-saccadic movement data. Thus, for clarity in the bivariate histograms, we have censored the DIEM saccade data to remove any events with a duration lasting for only a single sample (at the sampling rate of this means any sample less than 34 ms). We have not censored the non-saccadic movement data in the bivariate histogram plot. Figure 5 shows the maximum dispersion and maximum speed bivariate and univariate histograms for the two non-blink categories in the DIEM data.

From Fig. 5, the two categories exhibit similar ranges of maximum speed to each other, but the saccades in Fig. 5b exhibit a larger range of dispersion values. We note that this overlap seems to show the events tagged as saccades and fixations have speed estimates that do not match exactly with their respective definitions in the EyeLink data. This is

Fig. 5 Maximum dispersion–maximum speed histograms of the two non-blink categories in the open-source DIEM data. For clarity in these images, the depicted ranges of both dimensions are right-censored at the 0.95 quantile. See the text for details on data censoring and the correlation analysis results

likely because we are estimating speed in this figure based on the down-sampled data. However, despite the apparent overlap in the dispersion and speed ranges, the relationships between the two variables appear very different. For the (uncensored) non-saccadic movements in Fig. 5a, maximum speed and maximum dispersion exhibited a small but significant positive correlation, $r_{\text{NonSaccade}} = 0.08$ ($t(27,269) = 12.79$, $p < .001$), using a two-tailed paired sample t-test. For the censored saccade events in Fig. 5b, maximum dispersion and maximum speed exhibit a moderate correlation of $r_{\text{Saccade}} = 0.34$, ($t(7119) = 30.56$, $p < .001$), using a two-tailed paired sample t-test. Note that if we compute the correlation between maximum dispersion and maximum speed for the uncensored saccade data, the variables only exhibit a weak correlation, $r_{\text{Saccade}} = 0.17$, ($t(26,503) = 27.84$, $p < .001$), using a two-tailed paired sample t-test. This is because the set of saccades with a duration less than 34ms exhibited a mean maximum dispersion of 0° , which reduce the strength of the correlation with speed.

DIEM results

The BP-AR-HMM identified 7 unique states in the DIEM data. Like the Lund data, this was consistent across all 100 of the preserved MCMC samples. Of the 7 states, two (states 6 and 7) were associated with blinks, off-screen, or other errors and are collapsed into a single state in the confusion matrix in Table 4. Unlike the Lund data, events were more evenly distributed across the states, with 23%, 31%, 24%, and 12% of the samples in each of the four most frequent states.

From the total set (10 videos $\times$ 10 participants) BP-AR-HMM extracted 100,257 unique events. Out of the total events identified, 34,190 lasted for only one sample (33 ms).

Fig. 6 Example gaze trace from the DIEM dataset with color indicating BP-AR-HMM state. State 1 is dark blue; state 2 is blue; state 3 is red; state 4 is light red; and state 5 is yellow

Figure 6 shows sample speed and acceleration by time plots with color indicating state from the BP-AR-HMM modeling. Like the parse of the Lund data, the separate states match qualitative patterns expected from gaze path data: slow movements indicated in blue are connected by fast movements in shades of yellow and red in Fig. 6.

While we censored single-sample events in the Lund dataset, it is more reasonable that meaningful events could have occurred within 33 ms than 2 ms, so we have included these events in the descriptive statistics for the DIEM data in Table 3. However, we again right-censored the data based on acceleration outliers. For the DIEM data, because spuriously high acceleration samples were sparse and occurred in all states, the data were right-censored at the 90% quantile on acceleration within each category.

Table 3 summarizes the descriptive statistics for the five states in the DIEM data, computed on the right-censored data. Like the Lund data, the states are reasonably discriminated by the maximum dispersion and maximum speed of gaze movement (Table 3); again the plots of all dispersion, speed, and acceleration combinations are included in the supplemental material. Nonetheless, the distributions across states overlapped for both the maximum dispersion and, to a lesser degree, maximum speed. The states varied from the slowest, state 2, which exhibited maximum speed between 1.5 and 7.0 $^\circ/\text{s}$, to the fastest state 5, in which maximum

Table 3 Descriptive statistics for the BP-AR-HMM movement states on the DIEM data^a

	Mean	Std. Dev.	5% Quantile	95% Quantile
Maximum dispersion (deg.)				
State 1	0.21	0.38	0.00	0.98
State 2	0.27	0.23	0.00	0.69
State 3	0.96	1.28	0.00	3.27
State 4	2.70	2.71	0.00	7.03
State 5	8.55	3.56	0.21	13.36
Maximum speed (deg/sec)				
State 1	5.68	4.38	0.96	14.83
State 2	4.16	1.60	1.58	6.76
State 3	23.81	20.51	1.98	57.33
State 4	129.85	158.94	9.46	525.53
State 5	219.01	102.16	143.31	473.95
Maximum speed (deg/sec ²)				
State 1	74.79	43.26	16.52	160.28
State 2	60.28	20.41	26.53	93.34
State 3	309.73	253.56	42.96	913.15
State 4	1,249.84	1,299.86	282.13	4,503.57
State 5	1,899.05	899.76	1,175.30	4,237.96

^aNote that states 6 and 7 contain events with no meaningful speed or dispersion statistics

Fig. 7 Maximum speed-dispersion histograms for each BP-AR-HMM state in the DIEM data. For clarity in these images, the dispersion and speed axes are right-tail truncated at the 0.95 quantiles.

Sub-captions contain the Pearson correlation between maximum dispersion and maximum speed. See text for details on data censoring

speed ranged between 100 and 480 °/s. Thus, we again illustrate the BP-AR-HMM states by the bivariate and univariate maximum dispersion and maximum speed histograms in Fig. 7. The plots are illustrated with the axes truncated at the 95% for clarity, while the correlation analyses reported in the sub-captions are based on the full, right-censored data.

The slower states, 1 and 2, corresponded roughly with the slower states identified in the Lund set, although they were even slower on average in the DIEM dataset parsing. The bivariate histogram for state 2 (Fig. 7b) exhibits some similarity with state 1 (Fig. 3e) in the Lund data. Both have a soft upper bound on speed with more of the events within the state closer to that bound and the lower bound of zero maximum dispersion. State 1 seems to include a longer tail of both speed and dispersion behaviors compared to state 2.

Similarly, states 3, 4, and 5 (Fig. 7c, d, and e) corresponded roughly with state 3 from the Lund dataset. Most movements in the DIEM set were not as fast as those in states 4 and 5 in the Lund set. Interestingly, states 3, 4 and 5 appear to be bimodal distributions, perhaps each capturing two types of movement events. This bimodality perhaps accounts for the negative correlations observed in states 4 and 5.

Confusion analysis

Table 4 shows the confusion matrix of event assignments between the EyeLink tagging and the results of the BP-AR-HMM. Most fixation samples were in the slowest states (state 2, Fig. 7b and state 1, Fig. 7a), although the association was not as strong as in the Lund data. Indeed, a quarter of fixation samples were in state 3 and an additional 10% in state 4. Saccade samples were spread fairly evenly across the states, with more than 10% in each of the non-error states, although the majority were in states 3 and 4.

Following our analysis of the Lund data, we also examined the bivariate distributions of maximum dispersion by maximum speed when collapsing states 1 and 2 (collapsed grouping 1) as well as 3 and 4 (collapsed grouping 2). After collapsing, the number of single-sample events dropped to 2,393. Because collapsing removed much of noise in the dynamics, we did not censor the collapsed data for outliers.

We can see the three dominant trends in the relationship between maximum speed and maximum dispersion in

Fig. 8. Collapsed grouping 1 had the mean vector $<0.59^\circ, 19.61^\circ/\text{s}, 254.86^\circ/\text{s}^2>$ and standard deviation vector $<0.61^\circ, 66.72^\circ/\text{s}, 726.44^\circ/\text{s}^2>$. Collapsed grouping 2 had the mean vector $<3.84^\circ, 210.47^\circ/\text{s}, 1,899.48^\circ/\text{s}^2>$ and standard deviation vector $<3.24^\circ, 215.17^\circ/\text{s}, 1,972.46^\circ/\text{s}^2>$. Similar to the fixation-like state in the collapsed Lund analysis, there was only a small correlation in grouping 1. However, there was also a small negative correlation in the saccade-like collapsed grouping 2.

We note, however, that censoring the data in collapsed group 2 appears to remove the upper left cluster of data (low dispersion, high speed events). This is illustrated in Fig. 8c. This changes the central tendencies of collapsed grouping 2 to have mean vector $<0.27^\circ, 4.16^\circ/\text{s}, 60.27^\circ/\text{s}^2>$ and standard deviation vector $<0.23^\circ, 1.60^\circ/\text{s}, 20.44^\circ/\text{s}^2>$. The correlation now reflects a moderate positive relationship between dispersion and speed, more like the EyeLink saccade classification in the original DIEM data.

Discussion

We have demonstrated that by using an unsupervised approach a variety of gaze sequences could be described using very few states. These states aligned approximately with previously coded categories that are largely agreed upon in the eye-tracking literature. Thus, because of these findings and because of the flexibility of this approach to apply across sampling rates, we suggest that BP-AR-HMM constitutes an important initial parsing step that can be incorporated into any study as a first, exploratory processing step.

One notable difference between the results of the BP-AR-HMM and both the hand-coded data and the EyeLink parsed data was that the BP-AR-HMM produced more event types (states). Indeed, Andersson et al. (2016) has demonstrated that different approaches lead to significant differences in event numbers between parsing algorithms. Their finding and our results suggest that both human coders and existing parsing algorithms could be frequently underestimating the number and variety of events contained in a single eye-tracking recording. If a smaller number of unique events is desirable, researchers can use the BP-AR-HMM states to obtain an initial classification, parsing out major groups of events in the data. This can be followed by further analyzing the states for specific types of movements and interpretations grounded in the experimental design. Some potential additional parsing steps are discussed below.

Usage with additional parsing algorithms

A number of algorithms have been developed to support a second iteration of eye-movement parsing after periods of

Table 4 Confusion matrix of BP-AR-HMM states (columns) against EyeLink encoding (rows; Mital et al., 2011)

	State 1	State 2	State 3	State 4	State 5	State 6–7
Blink	0	0	0	0	0	1.0
Non-Saccade	0.265	0.361	0.249	0.105	0.020	0
Saccade	0.149	0.139	0.321	0.278	0.112	0

Fig. 8 **a–b** Maximum dispersion–speed histograms for the collapsed states 1 and 2 from the five non-blink DIEM states, showing similar bivariate distribution shapes but with different dispersion and speed ranges. **c** The maximum dispersion–speed histograms collapsed group two with the data censored for on the acceleration dimension at the 90% quantile. For clarity in all images, the ranges of both dimensions are right-truncated at the 0.95 quantile. Sub-captions contain the Pearson correlation between maximum dispersion and maximum speed

fixation and saccade activity have been separated. Most of these can operate agnostic of the particular fixation-saccade separation technique, and some simply rely on the basic fixed threshold methods. Such algorithms could easily be used to provide further parsing of the states found through the BP-AR-HMM method to further explore the characteristics of the data assigned to each state and to assign meaningful labels to those movement dynamics. Given visual evidence for bimodal distributions of dispersion-speed data within some categories, such as in Fig. 3a and c, further data parsing might elucidate if those states do contain more than one type of behavior with globally similar dynamics.

For example, the algorithms proposed by Engbert and colleagues, either with a static threshold (Engbert and

Kliegl, 2003) or an adaptive threshold (Engbert & Mergenthaler, 2006), were designed to decompose fixations into the component eye movements of tremor, drift, and microsaccade. Their approach begins by translating spatial coordinates of fixations into velocity by computing a 5-sample moving average velocity ± 2 points around each sample point. A microsaccade detection threshold is then defined as a function of the standard deviation of the velocity estimates; microsaccades are binocular events identifiable as relative outliers on the velocity distribution. Similarly, Otero-Millan, Castro, Macknik, and Martinez-Conde (2014) proposed a decomposition by using PCA on the multivariate characteristics of points of peak velocity within fixations. Clustering is used to tease out microsaccades, tremors, and

drifts. Critically, both algorithms are used after all fixations have been identified and blinks, errors, and other movements have been discarded. In the present data, for example, we might apply one of these algorithms to sequences of states 1 or 2 from the Lund data (Fig. 3d, e) or DIEM states 1–2 (Fig. 7a and b), wherein we see points with higher probabilities of being labeled fixations.

Additional analyses may further discriminate between fixations and smooth pursuit. For example, the metric for determining the degree to which a movement is in one direction from MarkEye (the ratio of PC scores for a window around a given time point) may be used on the state sequences identified by BP-AR-HMM. Larsson, Nyström, Andersson, and Stridh (2015) survey a number of other alternative metrics for post-processing that may be useful for discriminating fixation from pursuit. Likewise, the algorithm for parsing saccades and post-saccadic oscillations by Larsson et al. (2013) might be applied to states showing strong saccade-like characteristics Lund states 3 and 4 (Fig. 3b and c) or DIEM states 3 and 4 (Fig. 7c and d). The same algorithm may help parse the subgroups of movements indicated by the multi-modal bivariate distributions in the DIEM data.

In addition to further dividing states identified by BP-AR-HMM, it may also be desirable to consolidate particular sequences of states into a new unique state. For example, pursuit movements have a characteristic sequence of an initial ramp-up phase of high velocity, a second phase in which the velocity is relatively stable, and finally a ramp-down phase at the end of the pursuit. The BP-AR-HMM may classify the ramp-up and ramp-down as separate states from the stable phase. Hence, if a representation of a pursuit movement as a single state were desired, this pattern of states could be detected and consolidated.

Relationship to cognitive mechanisms

Multiple levels of analysis are important for interpreting the measurable eye-movement dynamics for the ultimate goal of understanding perceptual and cognitive processes manifest in those movements. Important distinctions must be made, for example, between movements that reflect active perceptual processing versus non-perceptual activity (Nakatani and van Leeuwen, 2007) or overt versus covert attention shifts (Engbert & Kliegl, 2003). Nuances in eye-movement dynamics can be discovered with the BP-AR-HMM in a way that is specific to the eye-tracker characteristics (e.g., sampling rate) and stimulus or task properties while being less dependent on the assumptions made by various pre-determined threshold algorithms. Thus, if an experiment is designed to elicit particular types of movements, the BP-AR-HMM approach could be used as an early, inexpensive (in terms of experimenter time and effort)

exploratory data analysis step to determine if data with the target characteristics have been collected. The number of unique categories and the differences in the characteristics of those categories gives a good indication of the sensitivity of a particular experiment design, including task, stimuli, and recording equipment, to various behaviors.

The BP-AR-HMM process, however, is mute when it comes to direct assessment of the cognitive mechanisms underlying eye movements. Whether a saccade resulted from endogenous or exogenous attention, for example, it will be grouped into the same category as all other saccades if they exhibit the same movement dynamics. Similarly, fixations due to target information processing or mind-wandering, for example, will be grouped together, if they exhibit the same dynamics of relatively stable, localized eye positions. However, as demonstrated by Salvucci and colleagues (Salvucci & Anderson, 1998; Salvucci, 2000) hidden Markov models are a powerful tool for connecting types of eye-movement events to cognitive mechanisms if used in conjunction with computational cognitive models such as ACT-R. Computational cognitive models can encode different cognitive strategies in ways that simulate predictions about the eye movements that would be generated by the use of each strategy for a given task (for more on how models simulate eye movements, the reader is referred to Nyamsuren & Taatgen, 2013; Salvucci, 2001). A number of studies have developed methods for determining candidate cognitive strategies for a given task (e.g., the “model spawner”; Zhang & Hornof, 2014). Following the I-HMM approach, cognitive strategies or states might be connected to the HMM stage of the BP-AR-HMM process. Or like human data, the simulated eye movements can be explored with the BP-AR-HMM as an unsupervised analysis technique that might help determine nuances in the data produced by different categories that might be missed with other parsing techniques. So although BP-AR-HMM does not help us make direct connections to the cognitive mechanisms of interest, the mathematical techniques it engages do lend themselves to integration with techniques that will aid this process.

Towards standardized evaluation approaches

Exploratory data analysis is often the first step in the statistical evaluation of a set of data. It can be critically important when working with large, multivariate time series data, for researchers to get a sense early in the analysis process of the data characteristics, including the types and directions of trends or the amount of noise contaminating the sample. The BP-AR-HMM represents a flexible, data-driven analysis tool that can facilitate exploration of the spatial and dynamic characteristics contained in a gaze data set. Thus, it can serve to facilitate more exploratory analysis of the

data prior to making assumptions about the types of movement events present in a data set or setting the threshold values that define those movements. In both our high and low sampling rate applications, we observed five non-error states of movement events. In both cases, at least four of these states had meaningful speed and dispersion characteristics to be scrutinized (the other states captured errors and blinks and undefined or ill-defined errors or events). We have observed that multiple states from the exploratory process might correspond to a single type of gaze event, or that a single exploratory state might capture nuances within movement types that can be teased apart with further analysis. By remaining agnostic to the exact definitions of event types, the BP-AR-HMM approach can suggest to researchers just how many different types of behavior may be present in a data set. It may also suggest to them which algorithms may be appropriate for parsing those events into the known taxonomies.

The introduction of a systematic, data-driven approach to exploratory analysis as part of the eye-tracking research process could serve to help standardize the ways in which algorithms are selected and applied in the field. Indeed, Andersson et al. (2016) were recently critical of the lack of standard approaches to both the development and selection of eye-tracking parsing algorithms. They raised six key challenges that should be considered in more depth regarding algorithm development, evaluation, and selection. The BP-AR-HMM algorithm, particularly if used as a first step in a more data-driven exploration of the dynamics captured by a set of eye-tracking recordings, can contribute to positive progress on many facets of these challenges.

1. There is no consensus on evaluation techniques. This includes a lack of standard metrics or statistics for evaluating a new algorithm alone and in comparison to other existing metrics. This makes it difficult to determine if reported differences were due to the algorithm quality/capability or were an effect of the method of evaluation. Early exploration does not solve the challenge of evaluation statistics, but it can provide a thorough characterization of the data to help better identify dimensions along which the data vary. These dimensions would offer a data-driven determination of the metrics that should be used to evaluate the performance of the parsing algorithms. For example, multi-modality is observed in some states in the DIEM data plots in Fig. 7, particularly on the dispersion dimension. Metrics on the dispersion, or joint metrics on dispersion and speed, would be appropriate for studying the contents of these data.
2. The eye-tracking research community lacks a theoretically grounded definition of an event. As we mention

in our Introduction, many researchers rely on strict oculomotor definitions for events, while others rely on stimulus or context information to aid in the definitions. For example, smooth pursuit is often defined as an event/state that only occurs in the presence of a moving stimulus. If humans can not agree on event definitions or they employ variable criteria, we can not expect computer algorithms to come to a consensus during automated parsing. Further, not all algorithms detect all types of events, so the labeling of data may vary simply as a function of algorithm choice. BP-AR-HMM can be set to use the same types of data dynamics, such as always using a combination of velocity and acceleration, across data types, sampling frequencies, recording technology, and stimulus content, so we can have a degree of confidence that the it will behave similarly across applications. In the style of a meta-analysis, we might look across multiple implementations of BP-AR-HMM for consistency in the types of data states extracted. These might corroborate existing definitions of events, particularly in terms of their defining dynamic properties. They might also point toward new types of events or variability in properties that must be integrated into our definitions.

3. Many algorithms require humans to set the parameters and are therefore subject to biases or assumptions of the users. This is particularly problematic in static threshold algorithms. BP-AR-HMM provides a systematic way to extract the statistics from a data set that could inform those choices of parameters. BP-AR-HMM is sensitive to variability across participants or over time, and that can be leveraged to inform variability in the choice of parameters that would adapt to individual differences but provide consistent interpretations. It also offers a way to enable the study of variability in eye-tracking data through the variability in the statistics and parameters that best capture the data across algorithms. As is often done in the study of mathematical models, the role that the parameters play in capturing variability in the data might offer a new method of algorithm evaluation not often used in eye-tracking analysis to date.
4. There are many published algorithms, but there are few direct comparisons between them. Andersson et al. (2016) are critical of the literature in that many papers emphasize showcasing their own algorithm over the comparison between algorithms. Indeed, we are likely guilty of this as well. There are a number of hurdles to thorough comparison, including sharing and transfer-ability of code between laboratories, the above issues with definitions of the meaningful events, and the ongoing challenge that not all algorithms try to parse the same events. The last of these hurdles, in

particular, means direct comparison may be meaningless, because a parser separating smooth pursuit from fixations would have little to offer in comparison to an algorithm targeting the parsing of large saccadic movements. BP-AR-HMM does not solve this problem, because it is not a method for comparing algorithms or the output of those algorithms. However, introducing a standard exploratory data analysis step with something like BP-AR-HMM can aid researchers in making informed choices of the parsing algorithms that would be appropriate to apply to the data set under scrutiny or that is being used to test a novel algorithm. For example, if no cluster in the BP-AR-HMM output contains data that would indicate smooth pursuit, then algorithms emphasizing smooth pursuit parsing would be inappropriate for the data, even if the parameters could be finessed into producing a reasonable result. Similarly, if a lab is developing a novel algorithm for parsing smooth pursuit but the exploratory analysis of test data set does not contain evidence of smooth pursuit, then researchers can determine early that a different evaluation data set is needed.

5. There is growing use of dynamic scenes and naturalistic tasks, but many standard parsing techniques do not use methods that consider smooth pursuit or other motion-related artifacts. A side consequence of this is that although researchers may be aware of the lack of smooth pursuit measurement, the degree to which this is problematic in data and the extent it pervades the literature are both unknown. However, the findings of Agtzidis, Startsev, and Dorr (2016b) and Agtzidis, Startsev, and Dorr (2016a) indicate that on certain tasks, observers spend over 12% of their time performing smooth pursuit, certainly a nontrivial amount. This strongly implies the need to develop accurate smooth pursuit classification in current and future parsing techniques. The impact is that we currently do not have an understanding or means of assessing which other events (saccades, fixations, etc.) are most affected by smooth pursuit and other motion events.

As we have seen in the present applications, BP-AR-HMM is capable of distinguishing states of behavior that would appear to be the same according to some event definitions (which justifies collapsing across some states), such as those in Fig. 3b and c in the Lund data. Yet, some of those same states, when considered in multivariate space, show evidence of multiple clusters within them, such as the bivariate plots in Figs. 3a, 7c and d. In addition to velocity and acceleration, higher-order dynamics, such as the jerk or snap, or binocular data, might be integrated into the BP-AR-HM algorithm to see how those states change. The flexibility

of the BP-AR-HMM to leverage multivariate time series inputs offers a way to begin to quantify how complex movement dynamics are manifest in eye-gaze recordings.

If available, additional data streams, such as head or body motion trackers, could be integrated into the BP-AR-HMM process to define categories of events that are a function of both eye movements and other sources of motion. In this way, we can scale our exploratory process and definitions of eye-movement events to more complex and naturalistic tasks and recording environments. This might, in turn, help to advance the development of parsing algorithms capable of working in more complex scenarios.

6. Human experts can easily converse about eye-movement events, reflecting some common intuition, but because the definitions remain fuzzy, the extent to which experts actually would agree about the classifications within data sets is unknown. The consequences of this in the way algorithms are defined, calibrated, and evaluated is also unknown. Andersson et al. (2016) did try to quantify the differences that two raters brought to bear on a common data set for their algorithm evaluations. Studies of human evaluations and encoding are expensive and time consuming, and thereby rare. We do not suggest that human expertise can be replaced by exploratory processes like BP-AR-HMM. Rather, working with exploratory data analysis steps can impact this challenge in two ways. First, it creates a statistically grounded way to track characteristics that drive eye-tracking event classifications without expert judgment. If published consistently, the data can drive our common understanding of eye-movement event definitions that might later be extracted for cross-study meta analyses. Second, adding an exploratory step to the data processing pipeline can help researchers check their assumptions and biases about event definitions and algorithm parameterization.

There is some degree to which we might consider the output of the BP-AR-HMM process as a ground truth for the comparison of eye-parsing algorithms or even between human expert coders. Because it uses unsupervised learning, BP-AR-HMM can be applied to a data stream with minimal effort on the part of the researcher to encode events of interest prior to running the BP-AR-HMM process. When used in an unsupervised manner, then the BP-AR-HMM can down-select segments of interest from the time series of gaze data. If a researcher is only interested in particular types of movement events, then s/he can focus on the data captured in the state that contains the characteristics of those events, saving that researcher time searching for

those events. In our present analyses, we have used hand-coding and the EyeLink algorithm as a form of ground truth for making cursory interpretations of the BP-AR-HMM state contents. These inferences can be used in the other direction to query the output of the coding or parsing. For example, in the study of human expertise, we can ask if human experts were able to distinguish as many event types as the computational algorithms. This may lend itself to identifying subtleties for further training of experts. Alternatively, the BP-AR-HMM provide a common truth for comparing the performance of multiple humans, without needing to assume that one of those experts is more correct than another. Rather, we can assess their agreement with each other and with the unsupervised categories. In short, unsupervised classification opens up a bi-directional conversation between different parsing approaches by adding a third, data-driven set of classes to the eye-tracking parsing solution space. That is, rather than assigning events to data regardless of the actual complexities of the particular data samples, we can allow the actual data values contained in eye-tracking samples to drive the properties of the events to which they are assigned.

Conclusions

We have introduced a data-driven, unsupervised learning approach for parsing events captured in eye-tracking data. This approach uses a combination of a beta process, vector auto-regressive processes, and a hidden Markov model to parse estimates of eye-movement dynamics into categories of unique events defined by the dynamics themselves. Our example applications leveraged estimates of velocity and acceleration to classify samples. The approach is flexible enough to run on any combination of position, velocity, acceleration, jerk, or other moments of the time series dynamics. We have demonstrated flexibility of this approach across both high- and low-resolution eye-tracking data. Results of the process constitute a strong exploratory step that can be easily integrated into the eye-movement event parsing pipeline to support a data-driven understanding of the characteristics in a data set or to serve as an initial data-parsing step that can support other algorithms targeted at parsing hard-to-distinguish event times (e.g., fixations versus smooth pursuit).

Author Notes This research was supported by the Air Force Office of Scientific Research, Grant FA9550-13-1-0087 and sponsored by the U.S. Department of Energy (DOE) through the Analysis in Motion Initiative at Pacific Northwest National Laboratory. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. Preliminary versions of this work were presented at the 2016 Meeting of the Society for

Mathematical Psychology and at the 2016 meeting of the Society for Computers in Psychology. PNNL-SA-123278.

References

- Agtzidis, I., Startsev, M., & Dorr, M. (2016a). In the pursuit of (ground) truth: A hand-labelling tool for eye movements recorded during dynamic scene viewing. In *Proceedings of the second workshop on visualizing eye tracking data (ETVIS)*.
- Agtzidis, I., Startsev, M., & Dorr, M. (2016b). Smooth pursuit detection based on multiple observers. In *Proceedings of the ninth biennial acm symposium on eye tracking research & applications* (pp. 303–306).
- Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2016). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, 49(2), 1–22.
- Bahill, A. T., & Troost, B. T. (1979). Types of saccadic eye movements. *Neurology*, 29(8), 1150–1152.
- Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1), 164–171.
- Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *Journal of Vision*, 9(5), 19–19.
- Berg, D. J., Shen, J., & Itti, L. (2009). The eyeMarkup eye movement analysis tool. Retrieved from <http://ilab.usc.edu/cgi-bin/secure/viewcvcs.cgi/trunk/saliency/matlab/Eye-Markup/>
- Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4), 455–470.
- Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, 43(9), 1035–1045.
- Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences*, 103(18), 7192–7197.
- Forney, G. D. (1973). The Viterbi algorithm. *Proceedings of the IEEE*, 61(3), 268–278.
- Fox, E. B. (2009). *Bayesian nonparametric learning of complex dynamical phenomena* (Ph.D. Thesis). Cambridge: MIT.
- Fox, E. B., Hughes, M., Sudderth, E., & Jordan, M. (2014). Joint modeling of multiple related time series via the beta process with application to motion capture segmentation. *Annals of Applied Statistics*, 8(3), 1281–1313.
- Fox, E. B., Sudderth, E. B., Jordan, M. I., & Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics* 5, 1020–1056. <http://www.jstor.org/stable/23024915>
- Gelman, A., & Loken, E. (2014). The statistical crisis in science data-dependent analysis - a garden of forking paths - explains why many statistically significant comparisons don't hold up. *American Scientist*, 102(6), 460.
- Goldberg, J. H., & Schryver, J. C. (1995). Eye-gaze-contingent control of the computer interface: Methodology and example for zoom detection. *Behavior Research Methods, Instruments, & Computers*, 27(3), 338–350.
- Griffiths, T. L., & Ghahramani, Z. (2011). The Indian buffet process: An introduction and review. *Journal of Machine Learning Research*, 12(Apr), 1185–1224.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188–194.
- Henderson, J. M. (2016). The DIEM project. Retrieved from <https://thediemproject.wordpress.com/>

- Hughes, M. C. (2016). Nonparametric Bayesian inference for sequential data. includes beta process hidden Markov model (BP-HMM). Implemented in Matlab. Retrieved from <http://michaelchughes.github.io/NPBayesHMM/>
- Hughes, M. C., Fox, E. B., & Sudderth, E. B. (2012). Effective split-merge Monte Carlo methods for nonparametric models of sequential data. In *Advances in neural information processing systems* (pp. 1295–1303).
- Kapoula, Z., Robinson, D., & Hain, T. (1986). Motion of the eye immediately after a saccade. *Experimental Brain Research*, 61(2), 386–394.
- Karpov, A., & Komogortsev, O. (2011). Automated classification and scoring of smooth pursuit eye movements in presence of fixations and saccades (Tech. Rep. No TR2011-11-23). HCI Lab, Texas State University.
- Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering*, 57(11), 2635–2645.
- Land, M. F., & Tatler, B. W. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. Oxford: Oxford University Press.
- Larsson, L., Nyström, M., Andersson, R., & Stridh, M. (2015). Detection of fixations and smooth pursuit movements in high-speed eye-tracking data. *Biomedical Signal Processing and Control*, 18, 145–152.
- Larsson, L., Nyström, M., & Stridh, M. (2013). Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit. *IEEE Transactions on Biomedical Engineering*, 60(9), 2484–2493.
- Mason, R. L. (1976). Digital computer estimation of eye fixations. *Behavior Research Methods & Instrumentation*, 8(2), 185–188.
- Mital, P. K., Smith, T. J., Hill, R., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3(1), 5–24.
- Nakatani, C., & van Leeuwen, C. (2007). A pragmatic approach to multi-modality and non-normality in fixation duration studies of cognitive processes. *Journal of Eye Movement Research*, 1(2), 1–12.
- Nyamsuren, E., & Taatgen, N. A. (2013). Pre-attentive and attentive vision module. *Cognitive Systems Research*, 24, 62–71.
- Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, 42(1), 188–204.
- Otero-Millan, J., Castro, J. L. A., Macknik, S. L., & Martinez-Conde, S. (2014). Unsupervised clustering method to detect microsaccades. *Journal of Vision*, 14(2), 18, 1–17.
- Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision*, 8(14), 21–21.
- R Development Core Team (2011). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org> (ISBN 3-900051-07-0).
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Salvucci, D. D. (2000). An interactive model-based environment for eye-movement protocol analysis and visualization. In *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 57–63).
- Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, 1(4), 201–220.
- Salvucci, D. D., & Anderson, J. R. (1998). Tracing eye movement protocols with cognitive process models. In *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 923–928). Hillsdale: Lawrence Erlbaum Associates, Inc.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 71–78).
- Sauter, D., Martin, B., Di Renzo, N., & Vomscheid, C. (1991). Analysis of eye tracking movements using innovations generated by a Kalman filter. *Medical and Biological Engineering and Computing*, 29(1), 63–69.
- Shic, F., Scassellati, B., & Chawarska, K. (2008). The incomplete fixation measure. In *Proceedings of the 2008 symposium on eye tracking research and applications* (pp. 111–114).
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366.
- Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8), 16–16.
- SR Research (2007). EyeLink user manual [Computer software manual]. Ottawa, ON.
- Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, 25(2), 137–142.
- van der Lans, R., Wedel, M., & Pieters, R. (2011). Defining eye-fixation sequences across individuals and tasks: The Binocular-Individual Threshold (BIT) algorithm. *Behavior Research Methods*, 43(1), 239–257.
- Widdel, H. (1984). Operational problems in analysing eye movements. *Advances in Psychology*, 22, 21–29.
- Zhang, Y., & Hornof, A. J. (2014). Understanding multitasking through parallelized strategy exploration and individualized cognitive modeling. In *Proceedings of the 32nd annual acm conference on human factors in computing systems* (pp. 3885–3894).